

Oda Yankısını Gidermek için Seyreklik Tabanlı bir Düzenleme A Dereverberation Formulation Based on Sparsity

Aziz Koçanaoğulları ve İlker Bayram

İstanbul Teknik Üniversitesi
azizkocana@gmail.com, ibayram@itu.edu.tr

Özetçe

Oda yankısının giderilmesi problemi, oda dürtü cevabının bilindiği durumlarda, kısa zamanlı-Fourier dönüşümünden (KZFD) faydalanarak, seyreklik tabanlı basit bir enküçüklenme problemi olarak düzenlenebilir. Fakat böyle bir düzenlemede, özellikle oda dürtü cevabının KZFD’de kullanılan pencereden uzun olduğu durumlarda, ortaya çıkan maliyet fonksiyonunu en küçükleme KZFD işlecinin barındırmak zorunda kalır. Bu da, enküçüklemenin hesapsal maliyetinin artmasına yol açar. Bu bildiride, her yinelemede KZFD işlecine ihtiyaç duymayan ve basit bir yinelemeli algoritmayla çözülebilecek bir düzenleme ortaya koyuyoruz. Ayrıca, düzenlemenin gerçek oda yankısı kullanıldığı durumdaki başarımını inceliyoruz.

Abstract

The dereverberation problem can be cast as a sparsity-based minimization problem, by making use of the short-time Fourier transform (STFT), provided that the room impulse response is known. However, in such a formulation, especially if the room impulse response is longer than the window employed in the STFT, minimization of the resulting cost function requires the STFT operator at each iteration. This in turn increases the computation cost of the minimization task. In this paper, we propose a formulation which can be solved by a simple iterative algorithm that does not require the use of the STFT operator. We also investigate the performance of the formulation in a scenario where a real room impulse response is used.

1. Giriş

Kapalı alanlarda yapılan kayıtlarda kaynaklar oda dürtü cevabından etkilenirler. Bu bildiride, elimizde oda dürtü cevabı olduğu halde oda yankılı kayıttan kaynak işaretini geri kestirme problemini ele alıyoruz. Kaynağın zaman-sıklık düzlemindeki seyrek görüntüsünden faydalanan bir yöntem öneriyoruz.

Özel olarak, H oda dürtü cevabıyla evrişim operatörü, x kaynak işareti ve u ise gürültü olduğu halde elimizdeki gözlem

$$y = Hx + u \quad (1)$$

olsun. x ’in kısa-zamanlı Fourier dönüşümü (KZFD) katsayılarının (X ile gösterilsin) seyrek olduğunu varsayalım. Verilen bir işareti KZFD katsayılarına götüren operatörü S

ile gösterelim. Bu durumda, x ’in seyrekleştirilmiş KZFD katsayılarını

$$\hat{X} = \arg \min_Z \frac{1}{2} \|y - HS^*Z\|_2^2 + \lambda \|Z\|_1 \quad (2)$$

olarak kestirilmesi [4]’de önerilmiştir. $\hat{X}$ elde edildikten sonra zaman alanındaki işaret $\hat{x} = S\hat{X}$ olarak elde edebiliriz. (2)’deki problemi çözmek için yinelemeli eşikleme algoritması (ISTA) kullanılabilir. ISTA’nın yinelemeleri

$$Z^{k+1} = T_{\alpha\lambda} \left(Z^k + \alpha (HS^*)^* (y - HS^*Z^k) \right) \quad (3)$$

şeklinde. Burada $\tau > 0$ için T_τ , eşiği τ olan köşegen bir eşikleme (soft-thresholding) operatörüdür. Özel olarak, $x \in \mathbb{C}$ için,

$$T_\tau(x) = \begin{cases} 0, & \text{eğer } |x| < \tau, \\ (1 - \tau/|x|)x & \text{eğer } |x| \geq \tau. \end{cases} \quad (4)$$

T_τ operatörünün işlemsel maliyeti S ve H ’ye göre düşüktür ve (7)’deki yinelemelerde asıl yük HS^* operatörünün gerçekleştirilmesindedir. Bu bildiride, işlemsel yükü düşürmek amacıyla S ve S^* operatörlerini barındırmayan bir algoritma öneriyoruz ve başarımını inceliyoruz. Bunun için

$$SH = HS \quad (5)$$

denkliğini (yaklaşık olarak) sağlayan bir $\mathcal{H}$ operatöründen faydalanacağız. Böyle bir operatör bulduğumuz durumda (2) yerine

$$\tilde{X} = \arg \min_Z \frac{1}{2} \|Y - \mathcal{H}Z\|_2^2 + \lambda \|Z\|_1 \quad (6)$$

şeklinde bir kestirim önerebiliriz (burada $Y = Sy$). Bu noktada, (2) ve (6)’daki problemlerin denk olmadığını not edelim. Bu problem için ISTA yinelemeleri

$$Z^{k+1} = T_{\alpha\lambda} \left(Z^k + \alpha \mathcal{H}^* (y - \mathcal{H}Z^k) \right) \quad (7)$$

şeklinde. $\mathcal{H}$ operatörünün yükü HS^* operatöründen düşük olduğu sürece bu yinelemeleri gerçekleştirmek işlemsel açıdan daha caziptir.

Oda dürtü cevabının uzunluğu KZFD’de kullanılan çerçeve uzunluğundan küçükse (5)’teki denkliği yaklaşık olarak sağlayan köşegen bir $\mathcal{H}$ operatörü bulabiliriz. Fakat tipik çerçeve boyları 30-60 ms aralığında ve oda dürtü cevapları birkaç yüz ms’ler mertebesinde olduğu için köşegen bir $\mathcal{H}$ operatörüyle yapılacak yaklaşım sınırlı olacaktır. Bu bildiride (5)’teki denkliği yaklaşık olarak sağlayan bir $\mathcal{H}$ operatörü bulmak için [5]’te önerilen tartışmadan faydalanacağız.

Bu tartışma, oda yankısı giderme problemi kapsamında, [5]'te tek bir mikrofon gözleminin bulunduğu durumlarda, [3]'te ise birden fazla mikrofon gözleminin bulunduğu durumlarda kullanılmıştır. Fakat her iki yöntem de işaretin KZFD katsayılarının seyrekliğini gözönünde bulundurmamaktadır. Seyreklik ön bilgisini kullanmanın enküçük kareler yöntemine göre avantajı Bölüm 3'te tartışılmaktadır.

2. Önerilen Düzenleme ve Çözümü

2.1. Notasyon

Bu bildiride KZFD'yi bir süzgeç dizisi olarak yorumlayacağız. Bunun için, $g(n)$ bir altgeçiren süzgeç olduğu halde, $g_k(n)$ süzgeçlerini

$$g_k(n) = g(n) \exp(-j k \Delta\omega n) \quad (8)$$

olarak tanımlayalım. Burada $\Delta\omega$, sıklık ekseninin nasıl örnekleneyeceğini belirleyen bir sabittir. Ayrıca N ile 'downsampling' operatörünü ($\downarrow N$) ile gösterelim. Bu durumda KZFD katsayılarını

$$x_k(n) = (\downarrow N) (x(n) * g_k(n)) \quad (9)$$

olarak tanımlıyoruz. Yukarıda da değindiğimiz gibi, x 'i X 'e götüren operatörü S ile göstereceğiz ve $S^*S = I$ olduğunu varsayacağız.

Ayrıca ayrık-zamanlı bir $x(n)$ işaretinin ayrık zamanlı Fourier dönüşümünü (AZFD), ' s ' verili bir sabit olduğu halde $\omega \in [s, s + 2\pi]$ için

$$X(\omega) = \sum_n x(n) e^{j\omega n} \quad (10)$$

olarak tanımlıyoruz. $X(\omega)$ 'nın diğer ω 'lardaki değeri, fonksiyonun 2π ile periyodik olduğu gözönünde bulundurularak hesaplanabilir.

2.2. Evrişimin KZFD katsayılarına etkisi

Bu bölümde işaretin bir süzgeçten geçirildiğinde KZFD katsayılarının nasıl etkilendiğini inceleyeceğiz. Bu incelemenin sonunda girişte de belirttiğimiz şekilde $HS = SH$ sağlayan bir $\mathcal{H}$ elde edeceğiz. Bu kapsamda daha detaylı bir tartışma için bkz. [5].

$y = h * x$ işaretinin k 'inci KZFD bantını düşünelim (yani $y_k(n)$ 'yi). $Y_k(\omega)$ 'nin basit bir ifadesini bulmak için g_k süzgecinin bant sınırlı olduğunu varsayalım. Özel olarak, ' s_k ', g_k 'nin merkez frekansı olduğu halde

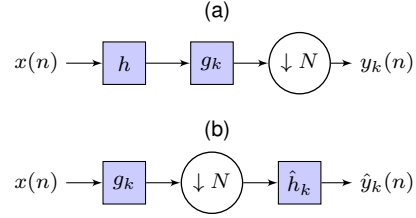
$$G_k(\omega) = 0, \quad \text{eğer } \omega \in \left[s_k - \pi, s_k - \frac{\pi}{N} \right] \cup \left[s_k + \frac{\pi}{N}, s_k + \pi \right] \quad (11)$$

olduğunu varsayalım. Bu durumda

$$Y_k(\omega) = X\left(\frac{\omega}{N}\right) H\left(\frac{\omega}{N}\right) G_k\left(\frac{\omega}{N}\right), \quad \text{eğer } \omega \in [N s_k - \pi, N s_k + \pi], \quad (12)$$

olduğu gösterilebilir. Bu noktada

$$\hat{H}_k(\omega) = H\left(\frac{\omega}{N}\right), \quad \text{eğer } \omega \in [N s_k - \pi, N s_k + \pi], \quad (13)$$



Şekil. 1: (a) y 'nin k inci KZFD altbantı bu sistemle hesaplanmaktadır. (b) Bant-sınırlılık koşulları sağlandığında (a)'daki sisteme denk başka bir sistem.

olarak tanımlarsak, Şekil 1b'deki sistemle elde edilen $\hat{y}_k, y_k$ 'ye denk olacaktır.

Pratikte kullandığımız g_k bant sınırlı değildir. Bu nedenle Şekil 1a,b'deki y_k ve $\hat{y}_k$ birbirinden farklı diziler olacaktır. Fakat g_k 'nin sıklık görüntüsünün s_k etrafında yeterince iyi yoğunlaşmış olduğu durumda $y_k \approx \hat{y}_k$ elde edebiliriz.

İlgili $\hat{h}_k$ süzgecini bulmak için

$$\min_U \int_{N s - \pi}^{N s + \pi} \left\| G_k\left(\frac{\omega}{N}\right) \left[H\left(\frac{\omega}{N}\right) - U(\omega) \right] \right\|_2^2 \quad (14)$$

probleminden faydalandık.

Bu problemi AZFD'nin bircil (unitary) olmasından faydalanarak zaman domeninde çözebiliriz. Bu şekilde $\hat{h}_k$ süzgecinin uzunluğunu da özgürce seçebiliriz.

Tipik olarak, uzunluğu L olan bir h süzgeci için $\lceil L/N \rceil$ uzunluğunda bir $\hat{h}_k$ süzgeci kullanarak iyi yaklaşımlar elde edebildiğimizi gördük.

Yukarıda anlattığımız anlattığımız yaklaştırmayı tüm KZFD bantları için tekrarlayarak $\hat{h}_k$ süzgeçlerini elde ettik. Bu süzgeçleri kullanarak, X verili bir x işaretinin KZFD katsayıları olduğu halde,

$$Y_k(n) = X_k(n) * \hat{h}_k(n) \quad (15)$$

şeklinde yeni bir çok kanallı işaret tanımlayalım. Bu durumda X 'i $\hat{Y}$ 'ye götüren operatörü $\hat{\mathcal{H}}$ ile göstereceğiz (yani $\hat{Y} = \hat{\mathcal{H}}X$). Bu durumda

$$\hat{Y} = \hat{\mathcal{H}}Sx \approx SHx = Y \quad (16)$$

olduğunu görebiliriz.

Yukarıdaki tartışmanın ne kadar güvenilir bir yaklaşım verdiğini test etmek için bir deneme yaptık. Bir konuşma işaretini 0.19 saniye uzunluğunda bir oda dürtü cevabıyla evriştirdik (frekans cevabı için bkz. Şekil 3b). Elde edilen oda yankılı işaretin KZFD bantlarından birinin gerçel kısmının bir parçası Şekil 2a'da görülebilir ('stem plot' olarak). Orijinal işaretin KZFD'sine $\hat{\mathcal{H}}$ uygulayarak elde ettiğimiz katsayıların gerçel kısımları ise yine aynı şekilde gösterilmiştir (boş halkalarla). Gözlenen altbantın genliği ve bahsi geçen iki (karmaşık değerli) bant arasındaki farkın mutlak değeri ise Şekil 2b'de gösterilmiştir. Fark işaretinin genliğinin göreceli olarak oldukça düşük olduğunu gözlemleyebiliriz.

2.3. KZFD Katsayıları Üzerinde Kestirim

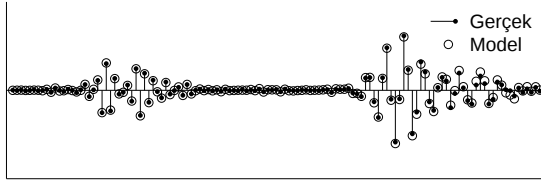
Bölüm 2.2'de anlatılan yaklaşım sayesinde

$$y(n) = h(n) * x(n) + u(n) \quad (17)$$

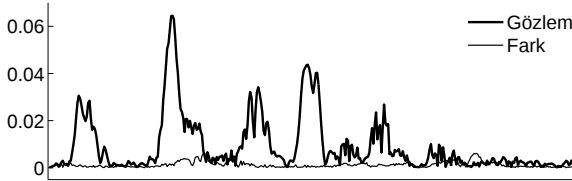
gözlem modeli yerine KZFD bantları üzerinde tanımlı

$$Y_k(n) = \hat{h}_k(n) * X_k(n) + \tilde{U}_k(n) \quad (18)$$

(a) Bir KZFD altbantından alınmış bir kesit (gerçek kısım)



(b) Bir KZFD altbantının ve yapılan hatanın genliği



Şekil 2: (a) Deneylerimizde kullandığımız konuşma işaretinin KZFD bantlarından birinden alınmış bir kesit ve bu kesitin Şekil 1b'deki sistem kullanıldığında elde edilen yaklaşımı. (b) (a)'da kısa bir kesiti verilen KZFD altbantının tamamının genliği (kalın çizgi) ve Şekil 1b'deki sistemle elde edilen yaklaşımın yaptığı hatanın genliği (ince çizgi).

gibi bir modeli kullanabiliriz. Burada, $\tilde{U}_k(n)$, sadece u 'nun KZFD'si değil, yaklaşım sonucu oluşan modeldeki hataları da barındıran bir gürültü işaretidir. (18)'deki modelin, (17)'deki modele göre daha cazip olmasının nedeni, $X_k(n)$ yaklaşık olarak seyrek olduğu halde $x(n)$ 'nin ancak KZFD operatörü uygulandığında seyrek bir gösterime sahip olmasıdır. Bunun sonucunda x 'i barındıran seyreklik tabanlı bir düzenleme için S gibi başka bir operatöre ihtiyaç duyarken, X_k için böyle bir operatöre ihtiyacımız yoktur. Bu çalışmada gözlenen işaretin KZFD katsayılarını (yani $Y_k(n)$) kullanarak X_k 'yi kestirmek için

$$\hat{X}_k(n) = \arg \min_z \frac{1}{2} \|Y_k(n) - \hat{h}_k(n) * z(n)\|_2^2 + \lambda \|z(n)\|_1 \quad (19)$$

gibi bir düzenleme kullandık. Bu dışbükey enküçükleme problemi yinelemeli eşikleme algoritmasıyla çözülebilir [2, 1]. Aşağıdaki algoritmada, $\hat{H}_k$ ve $\hat{H}_k^*$, sırasıyla $\hat{h}_k(n)$ ve $\hat{h}_k^*(-n)$ süzgeçleriyle evrişim işlemini göstermektedir.

Algoritma 1 Yinelemeli Eşikleme Algoritması

- 1: **tekrarla**
- 2: $\hat{X}_k(n) \leftarrow \hat{X}_k(n) + \alpha \left(\hat{H}_k^*(Y_k - \hat{H}_k \hat{X}_k) \right), \quad \forall k$
- 3: $X_k(n) \leftarrow T_{\lambda\alpha}(X_k(n)), \quad \forall k, n$
- 4: **yakınsayana kadar**
- 5: $\hat{x} \leftarrow S^* \hat{X}$

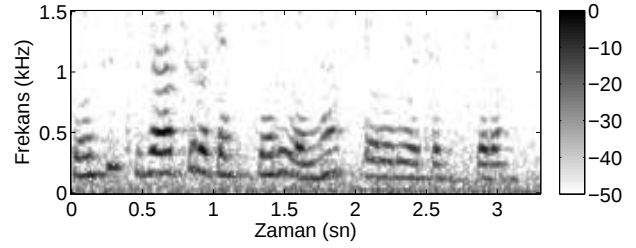
Bu algoritma, adım boyu α yeterince küçük alındığında yakınsamaktadır. Özel olarak,

Önerme 1. [1] $\sigma_k, \hat{H}_k^* \hat{H}_k$ operatörünün en yüksek öz değeri olsun. Tüm k 'ler için $\alpha \sigma < 2$ ise Algoritma 1, (19)'daki maliyet fonksiyonunun çözümlerinden birine yakınsar.

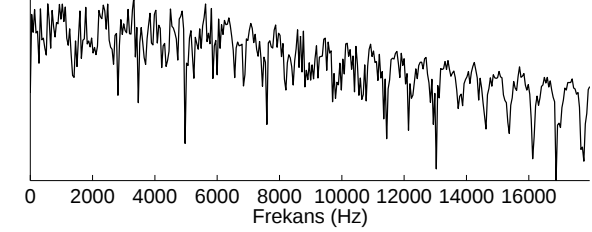
3. Deney ve Tartışma

Bildiride önerdiğimiz düzenlemenin başarımını anlamak için gerçek oda dürtü cevapları kullanarak müzik ve konuşma işaretleri üzerinde deneyler yaptık. Burada bir konuşma işareti kullanarak yaptığımız deneyi paylaşacağız.

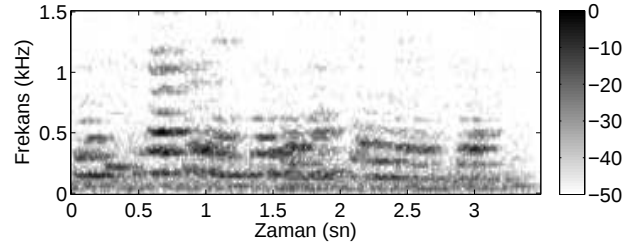
(a) Orijinal İşaret



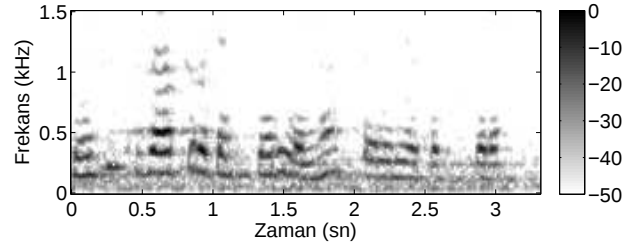
(b) Oda Frekans Cevabının Log-Genliği



(c) Gözlem İşareti



(d) Geri Çatılmış İşaret

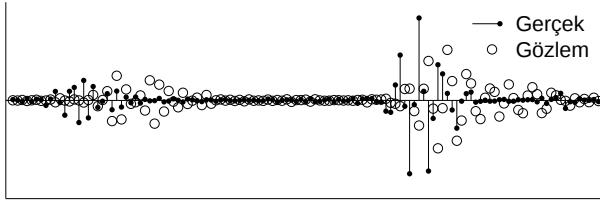


Şekil 3: Deneyde kullanılan işaretler. (a) Gürültüsüz konuşma işaretinin spektrogramı, (b) Oda dürtü cevabının genliğinin logaritması, (c) Orijinal işareti oda dürtü cevabıyla evriştirip bir miktar gürültü ekleyerek elde edilen gözlem işaretinin spektrogramı, (d) Gözlem işaretine önerilen yöntem uygulanarak geri çatılmış işaret.

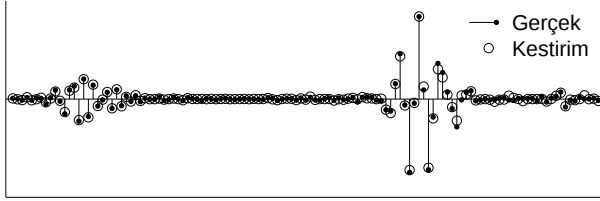
Deneyde spektrogramı Şekil 3a'da gösterilen bir konuşma işaretini (x) kullandık. Bu işareti 0.19 saniye uzunluğunda bir oda dürtü cevabıyla evriştirip bir miktar gürültü ekleyerek gözlem işaretini (y) elde ettik ($\text{İGO} = 20\text{dB}$). Oda sıklık cevabının genliği Şekil 3b'de gösterilmiştir. Ayrıca gözlem işaretinin spektrogramı Şekil 3c'de gösterilmiştir. Bu spektrogramda evrişimin etkilerini gözlemlemek mümkündür. Özellikle, Şekil 3a'daki doğuşkanların (harmonic) zaman eksenini boyunca uzadığını ve görece silikleştiğini görebiliriz. Bu etki, oda dürtü cevabının KZFD'de kullanılan pencerenin boyundan daha büyük olmasından kaynaklanmaktadır. Oda dürtü cevabı çok daha kısa olsaydı, KZFD'nin yaklaşık olarak bir maskeyle çarpılmış olmasını bekleyebilirdik.

Bu düzende, Algoritma 1'i kullanarak elde ettiğimiz geri çatılmış işaretin spektrogramı Şekil 3d'de gösterilmiştir. Geri çatılmış işaretin doğuşkanlarının orijinal işaretin doğuşkanları gibi kısa olduğunu not ediyoruz. Geri çatılan işarete oda yankısından kaynaklanan herhangi bir bozulma duyulmamak-

(a) Orijinal ve Gözlenen işaretlerin KZFD katsayılarından bir kesit



(b) Orijinal ve Geri Çatılmış işaretlerin KZFD katsayılarından bir kesit



Şekil 4: (a) Orijinal ve gözlenen işaretlerin KZFD altbantlarından birinden alınmış bir kesit. (b) Geri çatılmış işaretin aynı kesitteki KZFD katsayıları.

tadır.

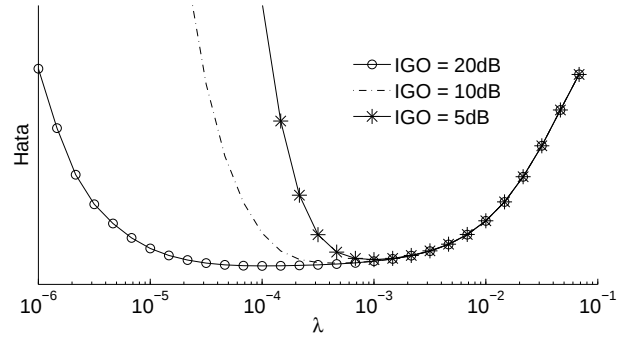
Gerçek çatılan işaretin KZFD katsayılarının orijinal işaretin KZFD katsayılarına ne kadar benzediğini görmek için enerjisi yüksek bir KZFD altbantını seçtik. Bu altbanttan bir kesit Şekil 4a'da gösterilmiştir (katsayıların gerçel kısımları kullanılmıştır). Katsayılarının birçoğunun sıfır veya sıfıra yakın olduğunu gözleyebiliriz. Bu gözlem bildiride kullanılan seyreklik varsayımını desteklemektedir. Aynı şekilde gözlenen işaretinin KZFD katsayıları da gösterilmiştir. Evrişim operatöründen dolayı bir miktar kayma ve sıfırdan farklı katsayıların sayısında artış gözlemlenebilir. Geri çatılan işaretin KZFD katsayıları ise Şekil 4b'de gösterilmiştir. Yöntem gerçek KZFD katsayılarını oldukça iyi bir şekilde kestirebilmiştir.

Son olarak, geri çatma yönteminin en küçük kareler yöntemine göre başarımda bir artış yaratıp yaratmadığını inceledik. Bunun için düzenlemedeki λ değerine bağlı olarak geri çatılan işaretin orijinal işaretten farkının enerjisini hesapladık. Aynı işlemi farklı işaret gürültü oranına (İGO) sahip gözlem işaretleri için tekrarladık – özellikle, İGO'nun 5, 10 ve 20 dB olduğu durumlarda elde ettiğimiz hata eğrileri Şekil 5'te gösterilmiştir. $\lambda = 0$ için elde edilen çözümün en küçük kareler yöntemiyle elde edilen çözümdür. İncelenen tüm İGO değerleri için hatayı en küçükleyen λ 'nın sıfırdan farklı olduğunu görüyoruz. Dolayısıyla 'seyreklik' ön bilgisinin aktif olarak kullanılmasının başarımla iyileştirdiğini söyleyebiliriz. Bu hata eğrilerine dair söylenecek bir başka nokta ise İGO düştükçe, en düşük hatayı veren λ değerinin arttığıdır. Fakat λ 'nın artması demek kullanılan düzenleyici terimin ağırlığının artması demektir – bu da ele alınan senaryoya tutarlıdır çünkü İGO düştükçe gözlenen işaretin güvenilirliği görece olarak düşmektedir.

4. Sonuç

Bu bildiride oda yankısı giderme problemi için hesap yükü düşük bir algoritmayla çözülebilen bir düzenleme önerdik. Düzenleme, KZFD operatörü ve evrişim operatörlerinin sırasının nasıl değiştirilebileceği tartışmasından faydalanmaktadır [5]. Düzenleme, elimizde tek bir mikrofondan yapılmış

Kestirilen işaretin λ 'ya göre değişen hata miktarı



Şekil 5: İGO = 5, 10 ve 20 dB olduğu durumlar için, düzenleyicinin ağırlığına, yani λ 'ya bağlı olarak geri çatılan işaretin gerçek işarete olan uzaklığı. $\lambda = 0$, en küçük kareler yöntemiyle elde edilen çözüme denk gelmektedir.

gözlemler olduğu durumlarda geçerli olsa da çok mikrofondan bir düzen için de kolaylıkla genelleştirilebilir. Ayrıca düzenlemede oda dürtü cevabının bilindiği varsayılmaktadır. Özellikle tek mikrofondan senaryolarda, oda dürtü cevabının bilinmediği durumda ortaya çıkan problem bir kör ters-evrişim (blind deconvolution) problemi olarak düşünülebilir. Birden fazla mikrofondan gözlem yapıldığında ise her bir mikrofona farklı bir oda dürtü cevabına sahip olacaktır. Fakat alta yatan işaretin seyrek olduğu ön bilgisi oda dürtü cevabını kestirmek için kullanılabilir (farklı bağlamdaki bir tartışma için bkz. [6]). İleriki çalışmalarımızda bu bildirideki düzenlemeyi çok mikrofondan ve oda dürtü cevabının bilinmediği bir senaryo için genelleştirmeyi planlıyoruz.

5. Kaynakça

- [1] P. L. Combettes ve V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Model. Simul.*, 4(4):1168–1200, November 2005.
- [2] M. A. T. Figueiredo ve R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Processing*, 12(8):906–916, August 2003.
- [3] N. D. Gaubitch ve P. A. Naylor. Subband inversion of multichannel acoustic systems. In P. A. Naylor and N. D. Gaubitch, editors, *Speech Dereverberation*. Springer, 2010.
- [4] M. Kowalski, E. Vincent, ve R. Gribonval. Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, 18(7):1818–1829, September 2010.
- [5] J.P. Reilly, M. Wilbur, M. Seibert, ve N. Ahmadvand. The complex subband decomposition and its application to the decimation of large adaptive filtering problems. *Signal Processing, IEEE Transactions on*, 50(11):2730–2743, Nov 2002.
- [6] F. Sroubek ve P. Milanfar. Robust multichannel blind deconvolution via fast alternating minimization. *IEEE Trans. Image Processing*, 21(4):1687–1700, April 2012.